DOCUMENT RESUME

ED 319 776 TM 014 994

AUTHOR Baker, Herbert George; And Others
TITLE Measuring Task Proficiency with Tailored Response Testing.
PUB DATE 90
NOTE 6p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; Evaluation Methods; *Fire Fighters; Job Performance; Job Skills; Military Personnel; Performance; Personnel Evaluation; *Simulation; *Test Format; *Videotape Cassettes; *Work Sample Tests
IDENTIFIERS Navy; *Tailored Response Testing

ABSTRACT

Tailored Response Testing (TRT) is a new type of test that has demonstrated its applicability to the evaluation of human performance in a wide variety of occupations and work settings. The Navy is using TRT to measure the technical proficiency of job incumbents in three of its jobs. The methodology holds great promise for testing aboard ships as well as at shore installations since it requires only a testing space containing a videocassette recorder. It is particularly promising where paper-and-pencil tests lack sufficient fidelity relative to the actual job tasks, but where job sample testing is not feasible, too costly, or dangerous. Results of the application of TRT to the Navy's fire controlman job are presented. Subjects were 103 first-term fire control personnel, whose scores ranged from 76 to 161, with a mean of 136.886 and a standard deviation of 18.16. Additional comparisons of TRTs with hands-on tests are planned. One table and one bar graph are included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. G. BAKER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED319776

MEASURING TASK PROFICIENCY WITH TAILORED RESPONSE TESTING

Herbert George Baker, Ph.D.
Navy Personnel Research and Development Center

Stanley J. Kalish, Jr., Ph.D.
Educational Testing Service

Laura E. Swirski
Navy Personnel Research and Development Center

ABSTRACT

Tailored Response Testing (TRT) is a new type of test that has demonstrated it applicability to the evaluation of human performance in a wide variety of occupations and work settings. The Navy is using TRT to measure the technical proficiency of job incumbents in three of its jobs. The TRT methodology holds great promise for test adminstration aboard ships as well as at shore installations, since it requires only a testing space containing a VCR. It is particularly promising where traditional paper-and-pencil tests lack sufficient fidelity relative to the actual job tasks, but where job sample testing is infeasible, dangerous, or too costly. This paper sketches out the concept of TRT and focuses on the application of TRT to measurement in the fire controlman job. Sample tasks are shown on video, as well as a practice TRT item and an actual test item. Descriptive statistics are presented for the tests administered in the Navy jobs.

INTRODUCTION

Tailored Response Testing

The Tailored Response Test (TRT) was developed by Richard F. Thornton of Educational Testing Service as an item type that permits an examinee to edit or tailor text so that the final edited, or tailored, text represents what the examinee believes are true statements. This assessment technique was intended to be a means of assessing performance in simulated siatuations, reflecting occupational situations. The simulation technique is intended to be of limited complexity so that developmental and operational costs can be within the range of smaller testing programs, i.e., 100 examinees or less, but can also be used in programs with a much larger examinee volume (Thornton, 1987).

A TRT is a set of sentences arranged in a meaningful format. Generally, the format will be sentences arranged in paragraphs, although other formats may be appropriate. The sentences may: (1) be entirely correct in terms of the subject content; (2) contain both accurate and incorrect

TM019994

BEST COPY AVAILABLE

subject content; or (3) contain completely inaccurate subject information.

Linking enlistment standards and job performance requires development and evaluation of performance measures and the comparison of performance on these measures with predictors used in selection and classification. However, excessive cost and adverse impact on operational organizations of hands-on tests would prohibit DoD-wide performance testing on the scale required for validity research. Therefore, a research strategy was evolved to (1) develop a hands-on test as a high-fidelity "benchmark" against which various surrogate measures could be compared, and (2) develop less expensive, easier-to-administer surrogate measures that could substitute for the hands-on measures. Research is currently focused on more than 30 occupational specialties across the four Services.

As a part of its contribution to the Joint-Service Project, the Navy is developing performance measures for a number of occupational specialties, or ratings, among which are the: (1) electronics technician (ET), (2) fire controlman (FC), and (3) gas turbine systems technician-mechanical (GSM). It is in the context of the search for surrogate measures that the TRT were constructed for each of those Navy ratings.

TRT FOR THE FC RATING

Critical Task Selection

A job analysis was conducted to identify critical tasks performed by first-termers, i.e., persons in their initial period of service, generally from 1 to 6 years. "Critical" means those tasks that: (1) are performed by a sizeable number of incumbents, (2) are important to mission success, (3) have at least moderate variance in performance, (4) are representative of the entire job domain to the greatest extent possible, and (5) are subsumed in the first-termers job across varied duty assignments. In addition, the tasks were restricted to those involving technical proficiency, which can be accomplished on an individual basis. The result was a set of job tasks comprising the technical domain of the first-term FC. From this task list, a subset of tasks, certified by subject matter experts (SME) constituting a comprehensive, yet safe and manageable testing package, was developed.

Hands-on Test Development

A hands-on test item was developed for each task contained in the final list of critical tasks. Each item is composed of observable, behavioral steps towards its completion. The behavioral steps for proper execution of the items are specified by SMEs, guided by manuals, written procedures,

Navy policies, and safety precautions. Each step must concur with policy, and performance on each step must be readily observable. The array of hands-on test itms for the FC is summarized in Table 1. (Note: The FC rating is split between data and radar subspecialties; therefore, there are actually two tests of equal difficulty for the FC rating.

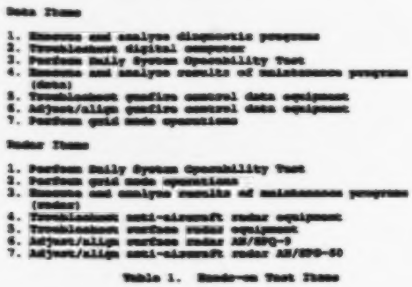


Table 1. Hands-on Test Items

Where materials are required in a test item, sample information and dummy classified materials were created. Specifically developed data input also ensured standardized test item content. Score sheets were developed for each test item, and contained a set of dichotomously scored elements corresponding to steps done correctly or incorrectly or to characteristics of task products that are acceptable or unacceptable.

TRT Test Development

The TRT was developed to match item-for-item with the hands-on test items. Each item was carefully scripted by the SME-Test Development team. Because hands-on tests require observable behavioral steps toward completion, they are generally suitable also for video-based TRTs. Nevertheless, the preparation for TRTs is very lengthy and exacting. An SME must precisely describe each and everything that must be done, what occurs (or may occur), what to observe, etc. Then, the myriad physical motions, dial settings, light sequences, and so forth are videotaped, often from several angles, and under varying conditions of illumination.

Subsequent to tape editing, voice-over narration was added as necessary. This was followed by preparation of instructional materials, answer booklets, and the videotape itself (augmented with practice items). The TRT was then pilot tested on a sample of FC job incumbents, refined, and readied for use in full-scale data collection.

TRT Administration

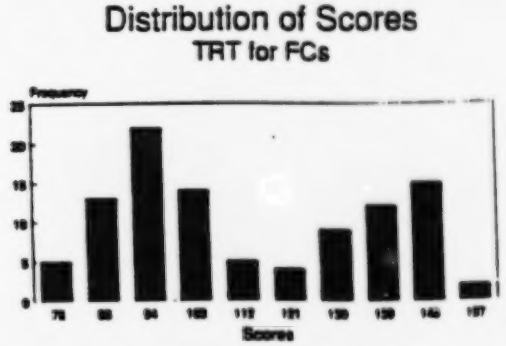
Subjects for the TRT were those who were being administered the hands-on tests. All subjects were drawn from active duty personnel who were first-termers in the rating.

Because TRTs incorporate on-screen directions as well as practice items, both TRT administrator training and testing session preparation time were minimal. Subjects were logged in, seated in front of the video cassette recorder (VCR) and given answer sheets, pencils, scratch paper, any necessary manuals, etc. The VCR was already loaded and ready to operate. The test subject pressed the "start" button, and proceeded according to directions given in the TRT itself. For each item, the subject "tailored" the response by crossing out words, phrases, and sentences as necessary, leaving the correct response.

Upon completion of the TRT, the test administrator collected the answer booklet and other materials, and readied the VCR for the next test subject. Answer sheets were scored using templates. Scores reflect points for correctly crossing out text elements, minus points for incorrectly crossing out text elements that should remain.

Results

Test results for the sample of FC TRT subjects (N=103) are presented in Table 2. The scores ranged from 76 to 161, a range of 85. The mean was 136.886, with a standard deviation of 18.16. The mode was 94, and the median score was 113.



It is apparent that the distribution of TRT scores exhibits bi-modality. It must be remembered that the TRT is netiher normed, nor pegged to any criterion other than critical tasks appropriate for first-term FCs.

CONCLUSION

Subsequent data analyses will assess the adequacy of the TRT as a surrogate for the hands-on tests by determining TRT score relationships to: (1) hands-on test global and item scores; and (2) aptitude test battery (ASVAB) composites as well as other predictors used in selection and classification.

The TRT research now underway constitutes the first application of this promising methodology to the technical occupations typical of the wide range of enlisted jobs in the Armed Services. Because of the large sample sizes and task-for-task tracking with job sample tests, important information will be gained relative to the characteristics, administrability, scoring, and general utility of Tailored Response Tests, particularly with technical jobs. This will contribute to the discipline of psychology in general and to testing theory and application in particular. If TRTs also prove to be adequate surrogates for hands-on tests, they could result in tremendous savings of money and time for performance measurement, and enhance personnel selection and classification in the Armed Services.

REFERENCES

Laabs, G.J., & Berry, V.M. (1987, August). The Navy Job Performance Measurement Program: Background, inception, and current status (NPRDC Tech. Rep. 87-34). San Diego: Navy Personnel Research and Development Center.

Thornton, R.F. Evaluation of a Scannable Method of Scoring Open-Ended Response Simulations. Proceedings of the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies, 1987, 119-130.